ME3023 Lecture - Chapter 4

Probability and Statistics

Theory and Design for Mechanical Measurements

5th ed. by Richard Figliola and Donald Beasley

Tristan Hill - Tennessee Technological University - Fall 2019

- **4.2 - Statistical Measurement Theory**

  - We want to estimate the true mean, x' from repeated measurement of x.

  - The true mean, x' is the average of all possible values of x. We never actually get this!

  - Through sampling we can find $\bar{x}$, the sample mean value of x. We do get this!

  - As our sample size increases, $\bar{x}$ approaches x'.

$$x' = \bar{x} \pm u_{\bar{x}}$$

- **Therefore, the sample mean $\bar{x}$ is the most probable estimate of the true mean x'.**

- **$\pm u_{\bar{x}}$ is the uncertainty interval in that estimate at some probability level, P%.**

- **The uncertainty interval is the range about $\bar{x}$ that you would expect x' to lie.**

- **4.3 - Statistical Measurement Theory**

The true variance is:

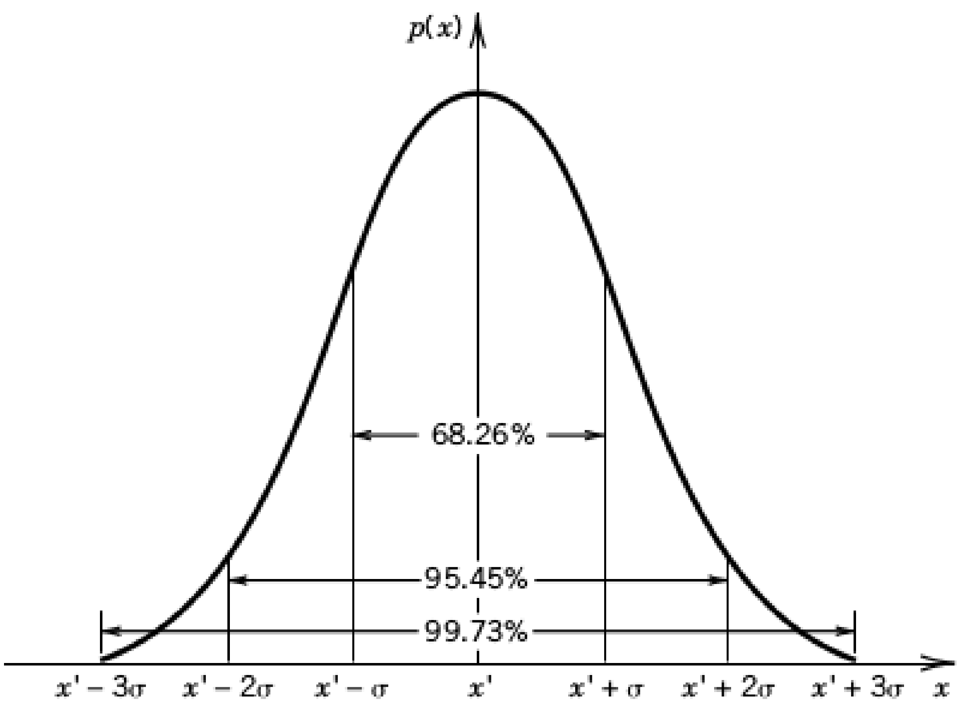
$$\sigma^2 = \int_{-\infty}^{\infty} (x - x')^2 p(x) dx$$

For discrete data the variance is:

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - x')^2$$

The standard deviation is the square root of the variance.

$$\sigma = \sqrt{\sigma^2}$$



- **4.4 - Statistics of Finite Sized Data Sets**

  - **We now try to predict the behavior of measured variable x based on a finite-sized sampling of x. We do this by comparing the statistics from that sampling to an assumed probability density function for the population. For example, if we recall the box of bearings discussed in Section 4.1, some two dozen bearings were measured, each having been randomly selected from a population numbering in the thousands. So how do we use the resulting statistics from this sampling to characterize the mean size and variance of all the bearings within the box? Within the constraints imposed by probability and if we assume a probability density function for the population, it is possible to estimate the true mean and true variance the population of all the bearings from the statistics of the sampling. The method is now discussed.**

  - **Suppose we examine the case where we obtain N measurements of x (that is, N repetitions), each measurement represented by x_i , where $i = 1, 2, ..., N$ and N is a finite value. In cases where N is not infinite or does not represent the total population, the statistical values calculated from such finite data sets are only estimates of the true statistics of the population of x. We will call such statistical estimates the finite statistics. An important point: whereas infinite statistics describe the true behavior of the population of a variable, finite statistics describe only the behavior of the sampled data set.**

– **Finite-sized data sets provide the statistical estimates known as:**

the sample mean:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

the Sample Variance:

$$s_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

The sample standard deviation is the square root of the sample variance.

$$s_x = \sqrt{s_x^2}$$

– **Did you notice that we divided by negative 1?**

– **The degrees of freedom, ν, in a statistical estimate equate to the number of data points minus the number of previously determined statistical parameters used in estimating that value. For example, the degrees of freedom in the sample variance is $\nu = N - 1$, as seen in denominator of the equations above.**

– **The relation between probability and infinite statistics can be extended to data sets of finite sample size with only some modification. When data sets are finite or smaller than the population, the** z **variable does not provide a reliable weight estimate of the true probability. However, the sample variance can be weighted in a similar manner so as to compensate for the difference between the finite statistical estimates and the statistics based on an assumed**$p(x)$**. For a normal distribution of** x **about some sample mean value,** x**, we can state that statistically**

$$x_i = \bar{x} \pm t_{\nu,P} s_x \quad (P\%)$$

$$t = \frac{\bar{x} - x'}{s_x / \sqrt{N}}$$

Table 4.4 Student's t Distribution

ν	t_{50}	t_{90}	t_{95}	t_{99}
1	1.000	6.314	12.706	63.657
2	0.816	2.920	4.303	9.925
3	0.765	2.353	3.182	5.841
4	0.741	2.132	2.770	4.604
5	0.727	2.015	2.571	4.032
6	0.718	1.943	2.447	3.707
7	0.711	1.895	2.365	3.499
8	0.706	1.860	2.306	3.355
9	0.703	1.833	2.262	3.250
10	0.700	1.812	2.228	3.169
11	0.697	1.796	2.201	3.106
12	0.695	1.782	2.179	3.055
13	0.694	1.771	2.160	3.012
14	0.692	1.761	2.145	2.977
15	0.691	1.753	2.131	2.947
16	0.690	1.746	2.120	2.921
17	0.689	1.740	2.110	2.898
18	0.688	1.734	2.101	2.878
19	0.688	1.729	2.093	2.861
20	0.687	1.725	2.086	2.845
21	0.686	1.721	2.080	2.831
30	0.683	1.697	2.042	2.750
40	0.681	1.684	2.021	2.704
50	0.680	1.679	2.010	2.679
60	0.679	1.671	2.000	2.660
∞	0.674	1.645	1.960	2.576

- **Why do we divide by N-1 when calculating the sample statistical parameters?**